REMARKS

Applicants and the undersigned are most grateful for the time and effort accorded the instant application by the Examiner. The Office is respectfully requested to reconsider the rejection present in the outstanding Office Action in light of the following remarks.

Claims 1-19 were pending in the instant application at the time of the outstanding Office Action. Of these claims, Claims 1, 10, and 19 are independent claims; the remaining claims are dependent claims. Claims 1, 10, and 19 have been rewritten. Applicants intend no change in the scope of the claims by the changes made by these amendments. It should also be noted these amendments are not in acquiescence of the Office's position on allowability of the claims, but merely to expedite prosecution. Further, it is respectfully asserted that these amendments to the claims find basis in the specification, specifically in the second paragraph of page 13.

The specification stands objected to because of a formula in the specification which is allegedly mathematically incorrect. However, Applicant would like to explain the formula so that the Office understands its correctness. An AST (atomic suffix tree) has a length of $n(n+1)/2$, wherein the length of the string S is n. Thus, where a string S has a length n (shown by the mathematical formula length(S)=n) the AST has a length of $n(n+1)/2$. Thus, as is aware, the formula is mathematically correct and applicable to the instant invention. Thus, reconsideration and withdrawal of this objection is respectfully requested.

Claims 1, 10, and 19 stand rejected under 35 USC § 112, first paragraph, as failing to comply with the written description requirement. Specifically, the claims contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors at the time the application was filed, has possession of the claimed invention. Applicants respectfully disagree that the latest Amendments to the Claims fail to comply with the written description requirement. Because the instant invention can be used for numerous languages, including Japanese, Chinese, and so forth, in which there is no word boundary in the language. Thus, it is inherent that the segmenting and splitting is necessarily not dependent upon word boundaries. Further, the specification explicitly asserts the methods utilized to segment and split the corpus into words, and utilizing word boundaries is not one of the asserted methods. Thus, reconsideration and withdrawal of this rejection is respectfully requested.

The same argument can be used to counter the 35 USC § 112, second paragraph, rejection. This type of segmenting and splitting of the instant invention can also be utilized for languages in which word boundaries exist. For example, in the English language, the following phrase "int hel ight" may produce the words "in", "the", and "light", regardless of the word boundaries that artificially produced the words "int", "hel", and "ight". Thus, reconsideration and withdrawal of this rejection is respectfully requested.

Claims 1-3, 6-12, and 15-19 stand rejected under 35 USC § 103(a) as being unpatentable over Wang et al. (hereinafter "Wang") in view of Razin et al. (hereinafter

"Razin") and further in view of Yang et al. (hereinafter "Yang"). Reconsideration and withdrawal of the present rejections are hereby respectfully requested.

The present invention is directed to a method and apparatus for automatically extracting new words from a cleaned corpus, where the corpus can be in any language that may or not have word boundaries (ranging from English or Latin to Chinese or Japanese). The instant invention segments a cleaned corpus to form a segmented corpus, splits the segmented corpus to form sub strings, and counts the occurrences of each sub strings appearing in the given corpus. Finally, the present invention filters out false candidates to output new words.

As best understood, Wang appears to be directed to a method that optimizes language models in which an initial language model is developed from a lexicon and segmentation derived from a received corpus. The initial model is iteratively refined by updating the lexicon and re-segmenting the corpus using both maximum match techniques and statistical principles. (Abstract) As asserted in the outstanding Office Action, Wang does not expressly disclose filtering out false candidates to output new words. Further, Wang does not expressly disclose that the segmenting and the splitting of the corpus is not dependent upon word boundaries. Nor does Wang disclose determining new words based upon the domain of the current corpus.

Razin fails to overcome the deficiencies of Wang as set forth above. As best understood, Razin appears to be directed to standardizing phrasing in a document. Razin identifies phrases in a document to create a preliminary list of phrases, then filters and

refines those phrases to create a final list of standard phrases. Razin then identifies phrase of a document that are similar to standard phrases, decides if the candidate phrase is similar enough to the standard phrase and compute phrase substitutions to determine the approximate conformation of the standard phrase to the approximate phrase and vice versa. (Abstract) There is no suggestion or teaching in Razin that the segmenting and the splitting of the corpus is not dependent upon word boundaries. In fact, Razin teaches away from this ability (column 11, lines 14-36), teaching that the source text is segmented using a standard finite-state machine technique that recognizes patterns that indicate word and sentence boundaries. Further, there is no suggestion or teaching that Razin discloses determining new words based upon the domain of the current corpus.

As best understood, Yang appears to be directed towards Chinese language modeling. Yang fails to overcome the deficiencies of Wang and Razin as asserted above. Specifically, Yang does not teach determining new words based upon the domain of the current corpus.

Claim 1 recites a "method of extracting new word automatically, said method comprising the steps of: segmenting a cleaned corpus to form a segmented corpus; splitting the segmented corpus to form sub strings, and counting the occurrences of each sub strings appearing in the corpus; and filtering out false candidates to output new words; wherein the segmenting and the splitting is not dependent upon word boundaries; **wherein new words may be determined based upon the domain of the cleaned corpus**". (emphasis added) Similar language also appears in the other Independent

Claims. Neither Wang, Razin, Yang, or any combination of the three, teach or suggest the limitations of the instant invention.

Further, a 35 USC 103(a) rejection requires that the combined cited references provide both the motivation to combine the references and an expectation of success. Not only is there no motivation to combine the references, no expectation of success, but actually combining the references would not produce the claimed invention. Thus, the claimed invention is patentable over the combined references and the state of the art.

Claims 4-5 and 13-14 stand rejected under 35 USC § 103(a) as being unpatentable over Wang et in view of Razin and Yang and further in view of Hui. Reconsideration and withdrawal of this rejection is hereby respectfully requested.

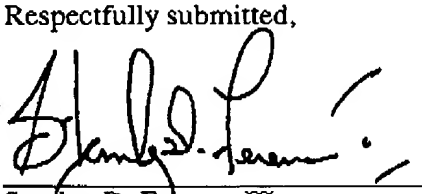
Hui does not overcome the deficiencies of Wang, Razin, or Yang. As best understood, Hui is directed towards an algorithm that provides an optimal sequential solution of the color set size problem which entails finding the number of different leaf colors in a subtree rooted at a vertex v in a rooted tree. Although Hui asserts that there is applicability in string matching heuristics, there is no teaching or suggestion in Hui that the segmenting and the splitting of the corpus is not dependent upon word boundaries or that new words can be determined based upon the domain of the current corpus.

Combining Wang, Razin, Yang, and Hui would result in producing a language model of phrases using an optimal sequential solution to find the phrases that constitute the lexicon of standard phrases. Even if there were a motivation for the combination, this combination does not teach or suggest the claimed invention.

In view of the foregoing, it is respectfully submitted that Independent Claims 1, 10 and 19 fully distinguish over the applied art and are thus allowable. By virtue of dependence from Claims 1 and 10, it is thus also submitted that Claims 2-9 and 11-18 are also allowable at this juncture.

In summary, it is respectfully submitted that the instant application, including Claims 1-19, is presently in condition for allowance. Notice to the effect is hereby earnestly solicited. If there are any further issues in this application, the Examiner is invited to contact the undersigned at the telephone number listed below.

Respectfully submitted,



Stanley D. Ference III
Registration No. 33,879

Customer No. 35195
FERENCE & ASSOCIATES
409 Broad Street
Pittsburgh, Pennsylvania 15143
(412) 741-8400
(412) 741-9292 - Facsimile

Attorneys for Applicants